UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*	2a. REPORT SECURITY CLASSIFICATION
USAMC Intern Training Center - USALMC Red River Army Depot ATTN: AMXMC-ITC-E-M Texarkana, Texas 75501	Unclassified
	2b. GROUP N/A

3. REPORT TITLE

THE GAMMA DISTRIBUTION IN DOWNTIME ANALYSIS.

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
N/A

5. AUTHOR(S) *(First name, middle initial, last name)*
Stephen M. Rolwing

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May, 1968	50	17

8a. CONTRACT OR GRANT NO. N/A	9a. ORIGINATOR'S REPORT NUMBER(S) N/A
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	N/A

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES N/A	12. SPONSORING MILITARY ACTIVITY Dir for Maint Hdqrs, USAMC Wash, D. C. 20315

13. ABSTRACT

This paper discusses the application of the three and four parameter Gamma distributions to problems involving downtimes. A method for estimating parameter values from empirical data is given for the three parameter distribution. The procedure is illustrated by several examples and a computer program capable of performing the required calculations is included.

In addition, this paper stresses the convenience of the Gamma distribution in downtime analysis. This distribution is especially useful in those cases in which convolutions are involved.
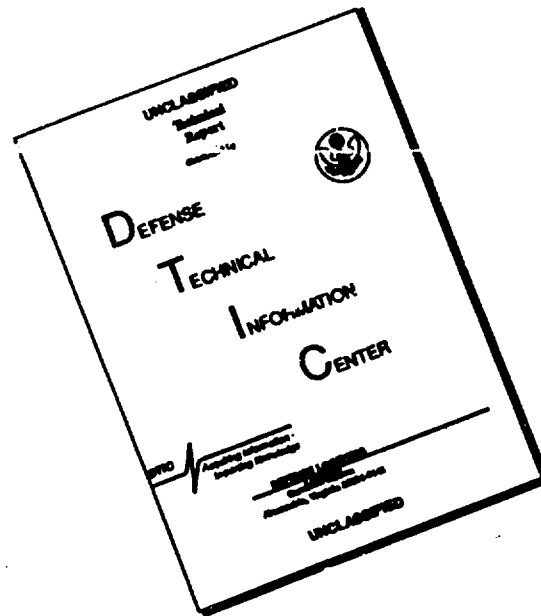
DD FORM 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unlimited
Security Classification

DISCLAIMER NOTICE

DEFENSE
TECHNICAL
INFORMATION
CENTER

THIS DOCUMENT IS BEST
QUALITY AVAILABLE. THE COPY
FURNISHED TO DTIC CONTAINED
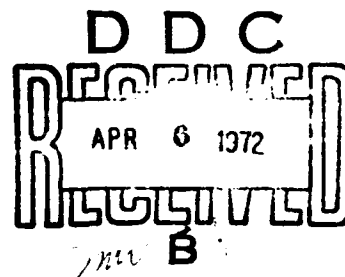A SIGNIFICANT NUMBER OF
PAGES WHICH DO NOT
REPRODUCE LEGIBLY.

RESEARCH REPORT

THE GAMMA DISTRIBUTION

IN DOWNTIME ANALYSIS

by

Stephen Michael Rolwing

Accomplished as Part of the
AMC Maintainability Engineering Graduate Program
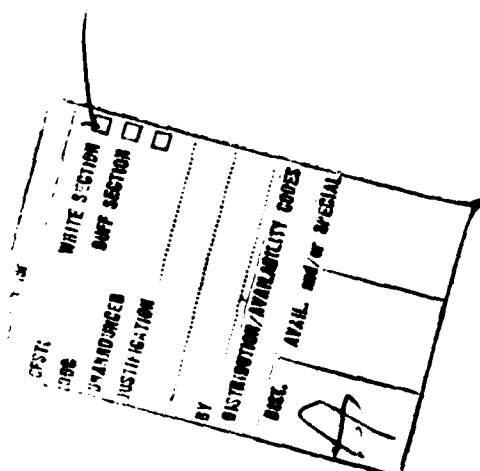USAMC Intern Training Center

and

Presented in Partial Fulfillment of the Requirements of the
Degree of Master of Engineering
Texas A&M University
May 1968

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

Maintainability ($\underline{M}$), as an engineering discipline, is
progressing through its infancy and is now gathering sophistication
and momentum. This growth can be seen in the increasing number of
published articles introducing new methods, new applications of old
methods, and new concepts slanted toward the field of maintainability.
This paper can be placed in the second category. It is an old
concept, applied in a different way to obtain the desired results.

The desired result in this case is the analysis of downtime
data by fitting a distribution to the data. The old method is the
familiar gamma distribution which has, in recent years, been the
subject of lively conversations as a possible distribution of
downtimes. It has been pointed out by many authors, Peterson (14)*,
in particular, that the gamma distribution can assume a variety of
different shapes. This is verified by Figure 1. In particular, note
that when the shape parameter is equal to 1, the distribution is an
exponential. Although the shape parameter is not restricted to an
integer value, an Erlang distribution (1) results when it is. As
the Erlang distribution has been used in queueing theory, some
further attributes of the gamma distribution may be found there.

Peterson (14), in his paper, discusses the fitting of a gamma
distribution to empirical data generated by a computer. Although
he discusses the attributes of the gamma distribution in the analysis

---

* Numbers in parentheses refer to the material listed in the
References at the end of this paper.
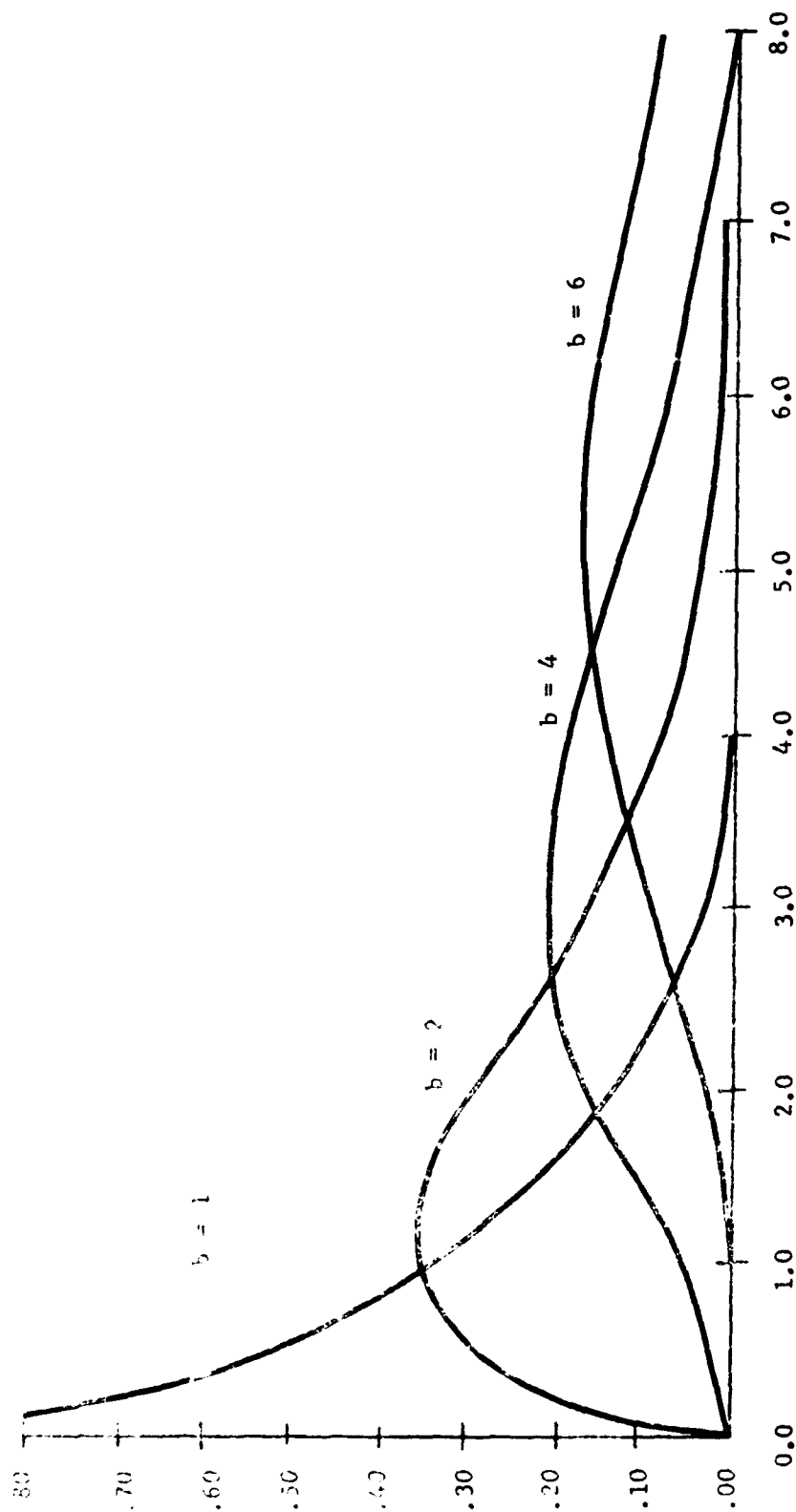
The Gamma Distribution for a=1, b=1, 2, 4, 6

FIGURE 1

2

of downtime to quite an extent, his method of parameter estimation
leaves much to be desired statistically. This method consisted of
plotting the empirical data and then "overlaying" this plot with
plots of the gamma distribution with integer shape parameters. The
gamma distribution which fitted best by this overlaying process was
then used to solve for the scale parameter. The threshold parameter
was assumed to be zero in all cases. Although the results were
biased by this method, he does give an actual application of the
gamma distribution fitted to downtime data.

The gamma distribution is attractive for two main reasons.
The first is that the log-normal distribution, which is usually
assumed to be the distribution of downtimes, is quite unwieldy and
requires plotting on log-normal paper to ascertain that the data is
actually log-normally distributed. The second reason is that many of
th tools and methods being brought into use today ..ometimes require
the manipulation of distributions in complex ways. For example,
Pieruschka (15) develops the force of ..ortality of equipment with
ideal repair and shows that a steady state value is obtained as the
number of failures increase. He uses Laplace transforms to obtain
the convolution of a gamma distribution with integer shape parameters
and then integrates the resulting distribution to find the mean.
These are the reasons why it is desirable to prove that the gamma
distribution will represent the anticipated data, and still be
mathematically simple.

Two downtime distributions can be spoken of in equipment analysis.
The first is the distribution of elapsed time required to complete

a specific maintenance task. The second concerns the distribution
of times required to complete different tasks within a given system,
and this distribution is said to describe the system downtime. The
random variable in the second case is, of course, composed of elements
of both cases, but the randomness of the individual task times is
usually neglected. This time element approach to the analysis of
downtimes will not be used here; but, instead, we will only be
interested in fitting a function to a number of data points.

The function to be fitted is the gamma distribution and para-
meter estimation is by maximum likelihood where certain terms of the
maximum likelihood equations have been tabulated. These tables are
entered with a parameter formed by the ratio of the arithmetic mean
to the logrithmic mean and the shape parameter term is read out.
Several more simple calculations will yield the threshold parameter
and the scale parameter.

In obtaining probability levels, the gamma distribution has one
of the disadvantages of the log-normal distribution in that tables
must be consulted (8,13). All of the moments of the gamma distribu-
tion are functions of the three parameters, and possibly most
important, a gamma distribution convolved with itself, results in
another gamma distribution.

To prove that the gamma distribution could be fitted to downtime
data, a computer program was written to perform the estimation pro-
cess. Random data from completely specified density functions
as well as empirical data were fitted with a gamma distribution. When
tested by the Kolmogorov-Smirnoff goodness of fit test (11), there

was no significant difference between any of the sets of data and its corresponding gamma distribution approximation at the .05 level.

A brief literature survey will be given in Chapter II, with the ensuing chapter devoted to the development and use of the gamma distribution and the estimation procedure. The convolution of the gamma distribution as pertains to downtime data is developed, and several examples will be given in Chapter IV. Some possible extensions of and conclusions to the paper will then be discussed in Chapter V.

CHAPTER II

LITERATURE REVIEW

Much of the impetus for this paper resulted from Peterson's
paper (14) which was a rough approximation of downtime data by the
gamma distribution. Initially, work was started on the so-called
generalized gamma distribution developed by Stacy (16) of the IBM
Corporation and at times worked on by Stacy and Mihram (17), Harter
(7), Harter and Moore (6), Parr and Webster (12), and Collins (3).
The fourth parameter is a power of the exponential term in the three
parameter gamma distribution. Although Stacy and Mihram and Collins
have developed two different parameter estimation procedures, the
methods are unwieldy and time consuming. The convolution of the
generalized gamma distribution is given as a summation term. In
attempting to fit the four parameter gamma distribution to data, it
was soon realized that the advantage of a wider variety of possible
shapes was not worth the additional complexity.

Although the method of moments in parameter estimation for the
gamma distribution is quite simple, the greater confidence obtained
by maximum likelihood is usually worthwhile. Several tables for
solving the maximum likelihood equations are available. Among these
are Chapman (2), Masuyama and Kuraiwa (1951), and Greenwood and
Durand (5), the last being used in this paper. General discussions
the gamma distribution are contained in most statistics and relia-
bility texts. Gupta (4) has developed the gamma distribution
to quite an extent in reliability analysis, and develops the

mortality and hazard rates of the distribution. Probability tables
for the Incomplete Gamma Function have been computed by Pearson (13)
and Harter (8). Some of these discussions are presented in the next
chapter in a general discussion of the gamma distribution.

CHAPTER III

THE GAMMA DISTRIBUTION

In this chapter, the gamma distribution will be investigated. The moments and the moment generating function along with the convolution of the gamma distribution will be given and the method of parameter estimation will be developed.

The gamma density function is given by

$$f(t) = \begin{cases} \dfrac{\left(\dfrac{t-c}{a}\right)^{b-1} e^{-\frac{t-c}{a}}}{a\ \Gamma(b)} & c < t < \infty \\ 0 & \text{otherwise} \end{cases}$$

where c is the threshold value,

   a is the scale parameter,

   b is the shape parameter, and

   $\Gamma(n)$ is the gamma function with argument n (Table 2).

The cumulative distribution is

$$F(t_0) = \Pr(t < t_0) = I(u,p)$$

where $I(u,p)$ is the Incomplete Gamma Function Ratio and in Harter's notation [8]

$$p = b - 1$$

and

$$u = \frac{t_0 - c}{a\sqrt{b-1}} \ .$$

The moments about the origin are found to be

mortality and hazard rates of the distribution. Probability tables for the Incomplete Gamma Function have been computed by Pearson (13) and Harter (8). Some of these discussions are preser ed in the next chapter in a general discussion of the gamma distribution.

$$\mu = a\,b + c$$

and

$$\sigma^2 = a^2 b.$$

In general the kth moment about the threshold parameter is

$$\mu_k = \frac{a^k \, \Gamma(b + k)}{\Gamma(b)} = a_k \prod_{j=0}^{k-1} (b + j).$$

A point that will be used later should be noted here, and that is that a number of different combinations of scale parameter and shape parameter will yield approximately the same distribution. This can be seen from the equations for the mean and variance in which there are three unknowns. An illustration of this point is shown in Figure 2 where the respective parameters of the curves are

Curve #	c	a	b	Mean	Variance
1	-.2	.65	4.15	2.5	1.75
2	0	.707	3.53	2.5	1.76
3	.4	.89	2.36	2.5	1.87

Parameter Estimation by Maximum Likelihood

In the development of the maximum likelihood equations for parameter estimation (5), we form the joint probability of drawing n samples, t_1, t_2, t_3...., t_n from a gamma distribution with parameters c, a, and b.

$$L = \left(\frac{(t_1-c)^{b-1} \, e^{-\frac{t_1-c}{a}}}{a^b \, \Gamma(b)} \right) \left(\frac{(t_2-c)^{b-1} \, e^{-\frac{t_2-c}{a}}}{a^b \, \Gamma(b)} \right) \ldots$$
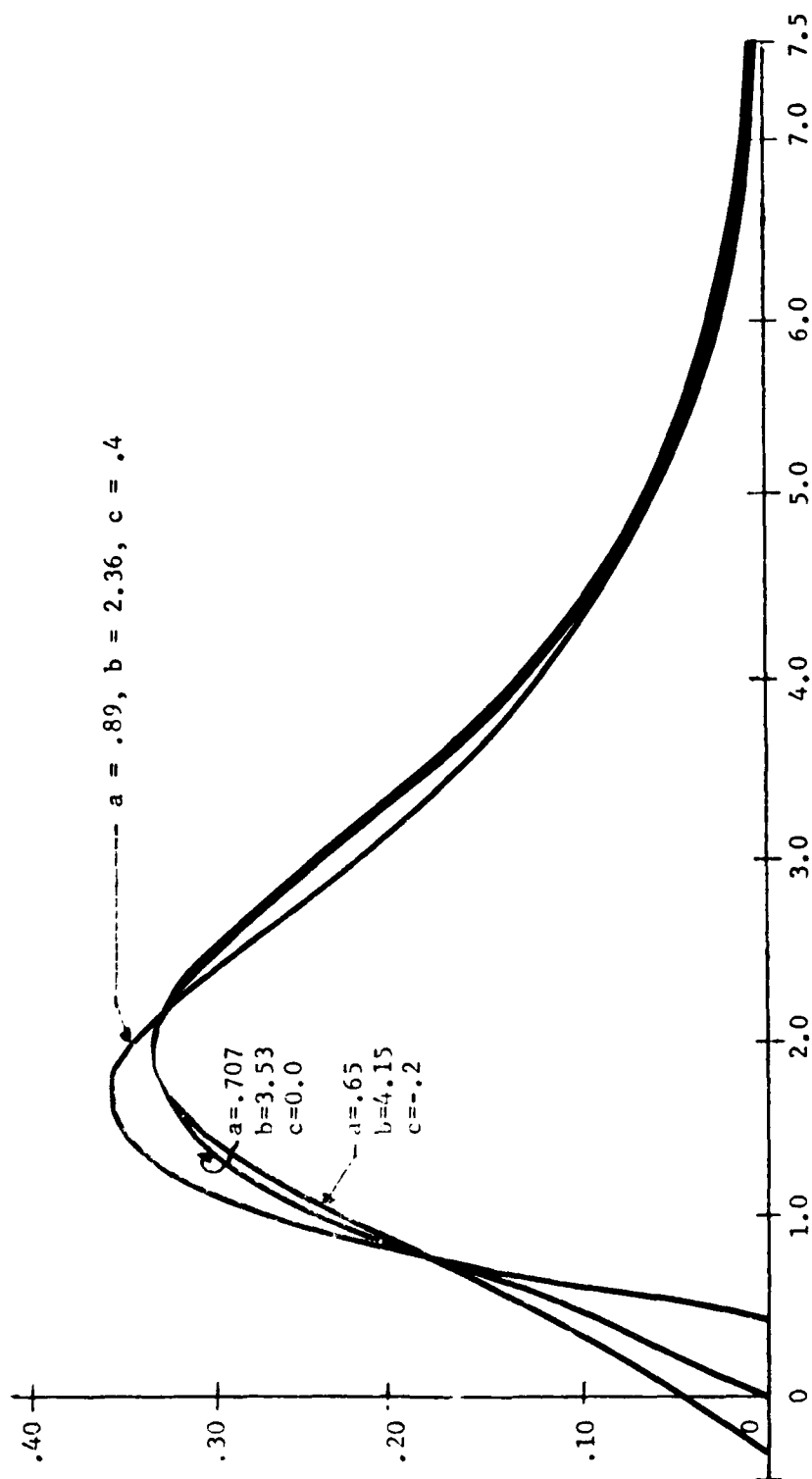
Gamma Distributions

FIGURE 2

a = .89, b = 2.36, c = .4

a=.707
b=3.53
c=0.0

a=.65
b=4.15
c=-.2

10

$$\left(\frac{(t_n - c)^{b-1} e^{-\frac{t_n - c}{a}}}{a^b \, \Gamma(b)} \right) = \frac{\prod_{i=1}^{n} (t_i - c)^{b-1} e^{-\sum_{i=1}^{n} \frac{t_i - c}{a}}}{a^{bn} \left[\Gamma(b) \right]^n} \; .$$

Taking the logarithm yields

$$\mathcal{L} = \ln L = (b-1) \sum_{i=1}^{n} \ln(t - c) - \sum_{i=1}^{n} \frac{t_i - c}{a} - bn \ln a - n \, \Gamma(b).$$

Differentiating with respect to the parameters to maximize the

probability of obtaining the given sample results in

$$\frac{\partial \mathcal{L}}{\partial a} = \sum_{i=1}^{n} \frac{t_i - c}{a^2} - \frac{nb}{a} = 0$$

$$ab = \frac{\Sigma t_i - c}{n} \qquad\qquad \text{Equation 1}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^{n} \ln(t_i - c) - n \ln a - n \frac{\Gamma'(b)}{\Gamma(b)} = 0.$$

Substituting the value found for a in equation 1, one obtains

$$\ln b - \frac{\Gamma'(b)}{\Gamma(b)} = \ln \frac{\Sigma (t_i - c)}{n} - \frac{\sum_{i=1}^{n} \ln(t_i - c)}{n} \qquad \text{Equation 2}$$

where $\Gamma'(b)$ is the first derivative of the gamma function with respect

to b. Taking the partial with respect to the threshold parameter gives

$$\frac{\partial \mathcal{L}}{\partial c} = - (b - 1) \, \Sigma \frac{1}{t_i - c} + \_ = 0$$

$$a (b-1) \sum_{i=1}^{n} \frac{1}{t_i - c} = n. \qquad\qquad \text{Equation 3}$$

In equation 1 the product ab is seen to be the **arithmetic mean**

corrected for the threshold value. In equation 2, $[\Gamma'(b)] \div [\Gamma(b)]$ is the tabulated digamma function (4), and equation 2 may be solved using a set of these tables and natural logarithms. However, the entire left-hand side of equation 2 has been tabulated in several different ways (2,5).

This then will be the method used to estimate the parameters. The right-hand side of equation 2 is simply the logarithm of the arithmetic mean minus the logarithm of the geometric mean. If we denote this right-hand side as Y and the left-hand side as $\eta(b)$, then

$$\eta(b) = Y.$$

Greenwood and Durand (5) have tabulated $b\eta(b) = b[\ln b - \Gamma'(b)/\Gamma(b)]$ versus $\eta(b) = Y$ to eight decimal places and their table is reproduced as Table 1.

The procedure will be to calculate the arithmetic and geometric means of the data and set Y equal to the difference of the logarithms of these two means. Enter Table 1 with this value and read $b\eta(b) = Yb$. Then solve for $b = Yb/Y$.

The scale parameter a is found from equation 1 and is equal to the arithmetic mean divided by b. Equation 3 indicates that some iteration process must be used to solve for the threshold value. Since this value must be found before the means can be calculated, a method of estimation which is free of terms containing a and b must be found. The following estimate was developed by fitting a straight line to the first three points on the curve generated by the expected ranks of the various ordered sample points. (Appendix 3 shows the development of this equation). The resulting expression is

$$c = \frac{Y_2 + Y_1}{2} - \frac{(Y_3 - Y_1)(Y_3 - Y_2)}{2(2Y_2 - Y_1 - Y_3)}$$

where Y_1, Y_2, and Y_3 are the first three ordered sample points.

Perhaps the most valuable asset of the gamma distribution is the ability to convolve one gamma distribution with another. The easiest proof of this is to derive the moment generating function and then find the distribution of the sum of random variables by multiplying their respective moment generating functions together. For the gamma distribution, the moment generating function is

$$E(e^{\theta t}) = \int_{-\infty}^{\infty} \frac{t^{b-1} e^{-\frac{t}{a}} e^{\theta t}}{a^b \Gamma(b)}$$

$$= \frac{1}{a^b \left(\frac{1}{a} - \theta\right)^b} \int_{-\infty}^{\infty} \frac{t^{b-1} \left(\frac{1}{a} - \theta\right)^b e^{-\left(\frac{1}{a} - \theta\right)t}}{\Gamma(b)} \, dt$$

$$= \frac{1}{(1 - a\theta)^b} \quad (1) \left(\frac{1}{1 - a^\theta}\right)^b$$

Thus, a gamma distribution convolved with itself will result in a gamma distribution with threshold parameter 2c, scale parameter a, and shape parameter 2b. In general, when n gamma distributions are convolved together, if they have a common scale parameter, a gamma distribution results with threshold parameter = $\sum_{i=1}^{n} c_i$, scale parameter a, and shape parameter = $\sum_{i=1}^{n} b_i$. The restriction of equal scale parameters for the distributions being convolved will not be too great of an impairment in the treatment of downtimes since each set of data will have approximately the same range. Recalling that a set of data

can be represented by a variety of scale and shape parameter combinations, it is seen that sets of data with unequal scale parameters can be convolved by preassigning a common scale parameter to each. The shape parameter for each set of data would then be found by dividing the individual arithmetic means by the common scale parameter. This concept will be discussed in the following chapter in the form of an example.

CHAPTER IV

EXAMPLES

The first example in this chapter shows the steps necessary to fit a gamma distribution to a set of data. The second example will illustrate the method of convolving several sets of data with unequal scale parameters.

1. The following times were taken in a maintenance test on a particular system. We wish to perform an analysis on the data and fit a distribution to it so that we may speak of the probability of completing a future action or actions within a certain time.

1.7	4.6	2.3	3.0
4.0	3.4	1.3	3.7
3.3	2.9	3.8	2.2
2.6	2.0	2.7	4.3
3.5	2.5	3.1	

The first step in the analysis is to find the first three ordered points. They are $Y_1 = 1.3$, $Y_2 = 1.7$, $Y_3 = 2.0$, then

$$c = \frac{1.7 + 1.3}{2} - \frac{(2.0 - 1.3)\ (2.0 - 1.7)}{2[(2)\ (1.7) - 1.3 - 2.0]}$$

$$= 1.5 - \frac{.21}{.2} = 1.5 - 1.05 = .45$$

The arithmetic mean is found to be

$$\bar{x} = \frac{\sum t_i - c}{n} \quad \frac{48.35}{19} = 2.542.$$

the geometric mean is

$$\text{antiln}\left(\frac{\sum\limits_{i=1}^{n} \ln (t_i - c)}{n}\right) = \text{antiln}\left(\frac{16.444}{19}\right) = 2.372,$$

then

$$Y = \ln\left(\frac{2.542}{2.372}\right) = .0694$$

Entering Table 1 we find that for

$$Y = .0694, \quad Yb = .5113.$$

Then

$$b = \frac{.5113}{.0694} = 7.37,$$

$$a = \frac{\bar{x}}{b} = \frac{2.54}{7.37} = .345,$$

$$\text{and } f(x) = \frac{(t - .45)^{6.37} \, e^{-\left(\frac{t - .45}{.345}\right)}}{.58}.$$

Entering the incomplete gamma function tables of Harter, where

$$p = b - 1 = 6.37 \approx 6.5$$

and

$$u = \frac{t - c}{a\sqrt{b}} = \frac{t - .45}{.935}$$

we can ascertain the cumulative probability for any given time and plot the graph of Figure 3. Also plotted on the graph is the curve from which the original data points were drawn, which in this case was a normal distribution with a mean of 3.0 and variance of 1.0. Performing the Kolmogorov-Smirnoff goodness of fit test (11)

$$D_{\text{MAX.}} = \text{MAX.} \left| F(t)_{\text{GAMMA}} - F(t)_{\text{NORMAL}} \right| = .06 < .301$$

where .301 is the critical value.

17

Fitted Gamma
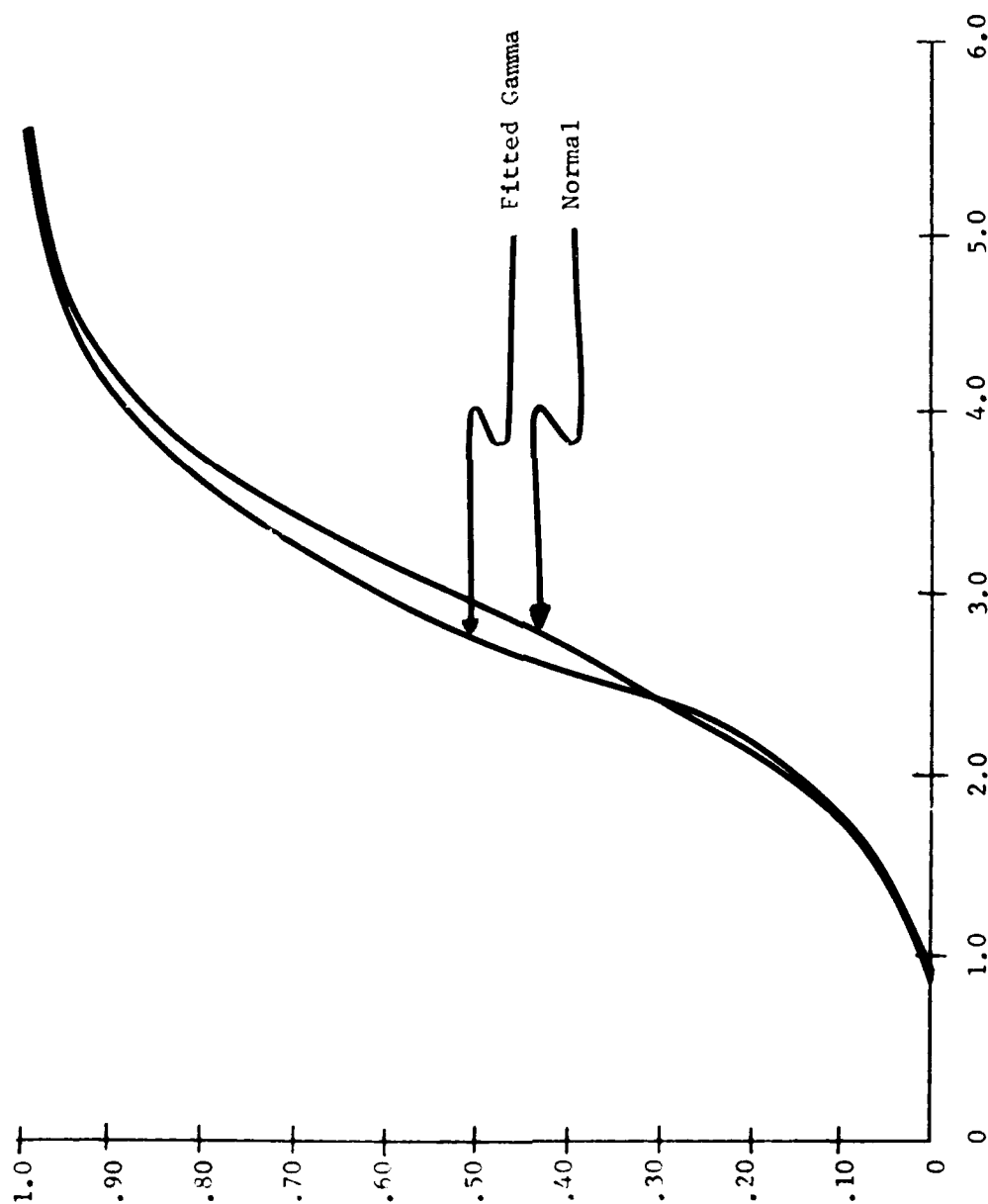
Normal

A GAMMA DISTRIBUTION FITTED TO NORMAL DATA

FIGURE 3

Therefore there is no significant difference at the .05 level of significance.

2.   A second example will demonstrate the convolution of a number of sets of data. The data is as follows:

System 1:

3.13	1.96	3.84	2.33
2.48	3.00	1.36	3.38
4.64	4.04	2.62	1.72
2.87	2.16	3.67	4.28
3.52	3.25	2.75	

System 2:

4.57	4.04	5.76	6.87
5.59	3.81	4.71	5.43
3.54	5.00	5.29	4.24
6.19	6.46	3.13	5.96
4.85	5.15	4.41	

System 3:

3.11	1.42	1.67	2.22
2.45	2.11	.58	.89
1.89	1.78	2.58	2.90
3.42	2.73	2.00	1.55
1.27	2.33	1.10	

System 4:

7.61	5.16	4.39	5.52
5.35	6.00	7.50	6.48
4.50	7.06	6.16	5.68
6.65	5.84	8.07	6.32
5.93	4.74	6.84	

Fitting the gamma distribution to the four systems as in the first example, the various parameters are found to be,

System	c	a	b	Mean	Variance
1	.94	.46	4.5	3.0	.95
2	2.68	.54	4.3	5.0	1.26
3	2.03	.34	11.7	6.0	1.35
4	.19	.39	4.7	2.0	.715

Setting $A = \dfrac{\sum_{i=1}^{4} A_i}{4} = .43,$

the b_i's are then

$$b_1 = \frac{\mu 1}{A} = 4.80,$$

$$b_2 = \frac{\mu 2}{A} = 5.45,$$

$$b_3 = \frac{\mu 3}{A} = 9.24, \text{ and}$$

$$b_4 = \frac{\mu 4}{A} = 4.21.$$

Then the distribution of the convolved systems is a gamma distribution with threshold parameter $C = \sum_{i=1}^{4} c_i = 5.84$, scale parameter $A = .43$, and shape parameter $B = \sum_{i=1}^{4} b_i = 23.7$. This distribution has been plotted in Figure 4 along with the convolution of the four data generating distributions which were all normal with means of 3, 5, 6 and 2 and variances of 1, 1.3, 1.6 and 7.5.

Note in particular the parameters of the total distribution

Method	Mean	Variance	$t_{95\%}$	$t_{99\%}$
Convolved Gamma	16	4.4	19.74	21.59
Convolved Normal	16	4.65	19.56	21.03

The Kolmogorov-Smirnoff goodness of fit was again applied with 4 x 19 = 76 degrees of freedom. No significant difference could be found at the .05 level.

CHAPTER V

CONCLUSIONS

Although the gamma distribution approximations found by the
author fitted extremely well, it is recommended that someone using
the method go slowly and check the initial results before proceeding.
This method will have important ramifications in the design of new
systems since the designer will know beforehand that the downtimes will
follow a gamma distribution of some sort. He can, therefore, feel free
to use the models necessary for proper design of a logistical system
without making a poorly based assumption as to the distribution of the
downtimes. Also, with the many time critical repair action situations
that occur in the space industry, a planning engineer can use the
accurate probability models without employing an inaccurate repair
time distribution.

The convolution technique developed for unequal scale parameters
needs more validation than the single example presented. However, the
potential of such a technique demands that it be investigated further.

For those who would still argue the validity of the log-normal
distribution in maintainability work, Figure 4 has been provided. It
is a gamma distribution fitted to a log-normal distribution with a
mean of .76 and variance of .604.

In the estimation procedure, a and b were fixed by the data as
soon as the threshold parameter was found, and varying c will vary
the fit. The equation for c on page 13 is certainly not the best
estimate. A possible alternative would be to plot the first few

Gamma (.5, .966, 2.07)

Log Normal (.76, .604)

$D_{max} = .02$

$n = 97$
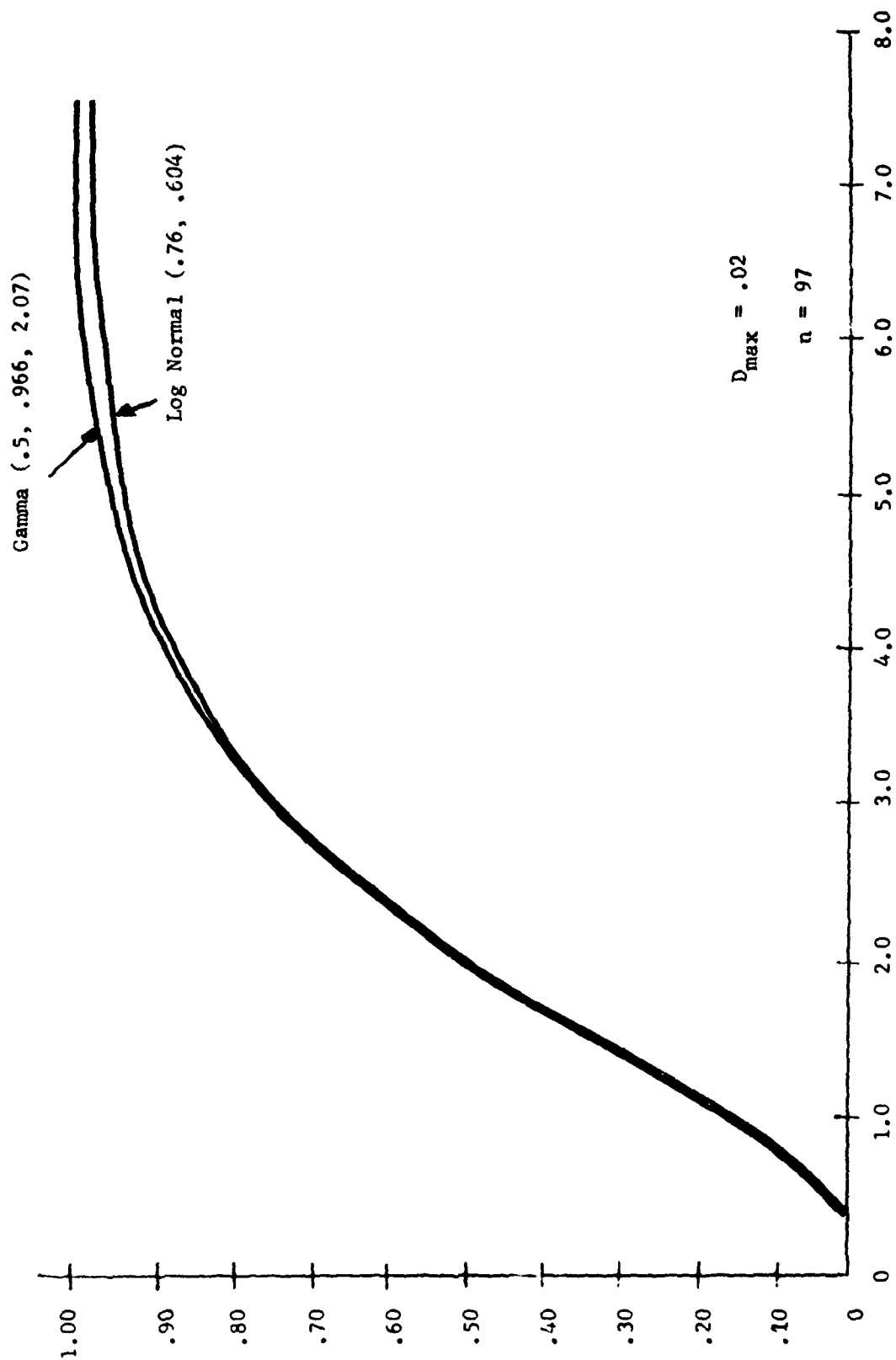
A Gamma Distribution Fitted to Log-Normal Data

FIGURE 4

sample points versus their expected, or median, ranks and approximating the threshold parameter by eye. This would certainly lend self-assurance to the estimate.

Although the process can be hand operated, computer programs have been written for both the data fitting process, and for the convolution technique and are presented with the expected output in Appendices 1 and 2.

Greenwood and Durand (5) have also developed the formulas for the variances of a and b. Since we violated one of their original premises by estimating c from the data, these formulas will not be presented in order to avoid temptation of the reader.

Further areas of research in the line of using the gamma distribution for representing repair times would be to improve the estimate of the threshold parameter in order to obtain a better fit. Perhaps this should be developed for a better fit in the upper half or perhaps only the upper tail. The convolution technique has assumed equal weight for all of the systems, this could most certainly not be the case. Further research could be done on weighting the different sets of data by their frequency with respect to the overall system. Interesting possibilities also lie in the previous research done on the Erlang distribution (1) which is a special case of the gamma distribution. Perhaps some of the theories which have been developed in queueing theory about the Erlang distribution would be useful in maintenance queueing analysis.

It is hoped by the author that further research on the gamma distribution will be done and that the gamma distribution will be used more frequently in downtime analysis. Since the gamma distribution is quite

simple in form, especially when the shape parameter is an integer, it is easy to become familiar with the distribution and gain an insight into the data it represents. It is hoped that this paper will help make the gamma distribution an available tool of the Maintainability engineer.

TABLE 1

Y VERSUS Yb

Y	Yb	Y	Yb	Y	Yb	Y	Yb
.00	.50000000	.21	.53239205	.41	.55846165	.61	.58061573
.01	.50166108	.22	.53380204	.42	.55965491	.62	.58163629
.02	.50331088	.23	.53520025	.43	.56083849	.63	.58264927
.04	.50494925	.24	.53658674	.44	.56201253	.64	.58365479
.04	.50657603	.25	.53796161	.45	.56317712	.65	.58465291
.05	.50819111						
.06	.50979438	.26	.53932495	.46	.56433238	.66	.58564375
.07	.51138573	.27	.54067685	.47	.56547842	.67	.58662738
.08	.51296508	.28	.54201741	.48	.56661536	.68	.58760389
.09	.51453237	.29	.54334674	.49	.56774331	.69	.58857337
.10	.51608755	.30	.54466493	.50	.56886236	.70	.58953591
.11	.51763057	.31	.54597209	.51	.56997264	.71	.59049158
.12	.51915141	.32	.54726833	.52	.57107426	.72	.59144048
.13	.52068005	.33	.54855377	.53	.57216730	.73	.59238267
.14	.52218649	.34	.54982850	.54	.57325189	.74	.59331825
.15	.52368074	.35	.55109265	.55	.57432812	.75	.59424729
.16	.52516282	.36	.55234633	.56	.57539610	.76	.59516987
.17	.52663277	.37	.55358965	.57	.57645593	.77	.59608607
.18	.52809061	.38	.55482273	.58	.57750771	.78	.59699595
.19	.52953640	.39	.55604568	.59	.57855154	.79	.59789961
.20	.53097019	.40	.55725862	.60	.57958751	.80	.59879710

TABLE 1 - Continued

.81	.5995885?	.96	.61237481	1.11	.62391639	1.26	.63448370
.82	.60057389	.97	.61317777	1.12	.62464951	1.27	.63515706
.83	.60145333	.98	.61397571	1.13	.62537836	1.28	.63582675
.84	.60232689	.99	.61476869	1.14	.62610298	1.29	.63649279
.85	.60219454	1.00	.61555677	1.15	.62682342	1.30	.63715523
.86	.60405665	1.01	.61633999	1.16	.62753971	1.31	.63781410
.87	.60491298	1.02	.61711841	1.17	.62825190	1.32	.63846943
.88	.60576369	1.03	.61789208	1.18	.62896003	1.33	.63912126
.89	.60660886	1.04	.61866105	1.19	.62966414	1.34	.63976962
.90	.60744854	1.05	.61942537	1.20	.63036427	1.35	.64041454
.91	.60828279	1.06	.62018509	1.21	.63106046	1.36	.64105606
.92	.60911168	1.07	.62094026	1.22	.63175276	1.37	.64169420
.93	.60993527	1.08	.62159093	1.23	.63244119	1.38	.64232900
.94	.61075362	1.09	.62243714	1.24	.63312580	1.39	.64296049
.95	.61156678	1.10	.62317895	1.25	.63380662	1.40	.64358870

TABLE 1 - Continued

Y	Yb	Y	Yb	Y	Yb
1.4	.6435887	7.2	.8010527	13.2	.8536588
1.6	.6555077	7.4	.8035935	13.4	.8548549
1.8	.6663340	7.6	.8060539	13.6	.8560277
2.0	.6723335	7.8	.8084382	13.8	.8571778
2.2	.6853382	8.0	.8107500	14.0	.8583059
2.4	.6937540	8.2	.8129931	14.2	.8594127
2.6	.7015677	8.4	.8151707	14.4	.8604989
2.8	.7088511	8.6	.8172861	14.6	.8615651
3.0	.7156639	8.8	.8193420	14.8	.8626118
3.2	.7220568	9.0	.8213412	15.0	.8636397
3.4	.7280727	9.2	.8232862	15.2	.8646493
3.6	.7337487	9.4	.8251795	15.4	.8656411
3.8	.7391165	9.6	.8270232	15.6	.8666156
4.0	.7442040	9.8	.8288196	15.8	.8675733
4.2	.7490353	10.0	.8305705	16.0	.8685146
4.4	.7536319	10.2	.8322778	16.2	.8694401
4.6	.7580126	10.4	.8339433	16.4	.8703402
4.8	.7621942	10.6	.8355687	16.6	.8712452
5.0	.7661916	10.8	.8371555	16.8	.8721255
5.2	.7700182	11.0	.8387052	17.0	.8729916
5.4	.7736861	11.2	.8402193	17.2	.8738438
5.6	.7772060	11.4	.8416989	17.4	.8746825
5.8	.7805879	11.6	.8431455	17.6	.8755080
6.0	.7838406	11.8	.8445602	17.8	.8763207
6.2	.7869722	12.0	.8459441	18.0	.8771208
6.4	.7899901	12.2	.8472983		
6.6	.7929012	12.4	.8486239		
6.8	.7957116	12.6	.8499219		
7.0	.7984270	12.8	.8511931		
		13.0	.8524384		

TABLE 2

GAMMA FUNCTION Γ(n)

n	0	1	2	3	4	5	6	7	8	9
1.0	1.0000	.9943	.9888	.9835	.9784	.9735	.9687	.9642	.9597	.9555
1.1	.9514	.9474	.9436	.9399	.9364	.9330	.9298	.9267	.9237	.9209
1.2	.9182	.9156	.9131	.9108	.9085	.9064	.9044	.9025	.9007	.8990
1.3	.8975	.8960	.8946	.8934	.8922	.8912	.8902	.8893	.8885	.8879
1.4	.8873	.8868	.8864	.8860	.8858	.8857	.8856	.8856	.8857	.8859
1.5	.8862	.8866	.8870	.8876	.8882	.8889	.8896	.8905	.8914	.8924
1.6	.8935	.8947	.8959	.8972	.8986	.9001	.9017	.9033	.9050	.9068
1.7	.9086	.9106	.9126	.9147	.9168	.9191	.9214	.9238	.9262	.9288
1.8	.9314	.9341	.9368	.9397	.9426	.9456	.9487	.9518	.9551	.9584
1.9	.9618	.9652	.9688	.9724	.9761	.9799	.9837	.9877	.9917	.9958

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x}\, dx. \qquad \Gamma(n) = (n-1) \cdot \Gamma(n-1),\ n > 1.$$

$$\Gamma(n) = (n-1)! \quad \text{when } n \text{ is a positive integer.}$$

APPENDIX 1

Computer Program for Fitting a Gamma
Distribution to a Set of Data

This appendix contains the flowchart, the program, and the
expected output of a program which fits a gamma distribution to n
data points. In addition, the input formats are given on standard
forms. The process is as in the main portion of this paper with the
following exceptions.

1. In lieu of Table 1, the following equations (5) were used to
solve for b.

$$b = y^{-1}(.5000876 + .1648852y - .0544274y^2)$$

For $0 < y < .5772$

$$b = \frac{8.898919 + 9.059950y + .9775373y^2}{y(17.79728 + 11.968477y + y^2)}$$

For $.5772 < y < 17.0$

2. The geometric mean was calculated by

$$\bar{t}_G = \prod_{i=1}^{n} t_i^{\frac{1}{n}}$$

FLOWCHART

Read the Number of Sample Points

Read the Sample Values

$\Sigma \; x_i = 0$

$\Pi \; x_i = 1.0$

Find the Threshold Parameter c

DO I = 1, N

$z_i = x_i - c$

$\Sigma \; z_i = \Sigma \; z_i + z_i$

$\Pi \; z_i = (\Pi \; z_i) \; (z_i)$

$\bar{x} = \dfrac{\Sigma \; z_i}{N}$

1

①

$$\bar{x}_G = (\Pi\ z_i)^{1/N}$$

$$Y = \ln\ (\bar{x}) - \ln\ (\bar{x}_G)$$

If
Y-.5772

Find Scale Parameter B

Find Scale Parameter B

Scale Parameter a = $\dfrac{X}{b}$

Find $\Gamma(b)$

DO I = 1, N

Find the Density Function
at Each Sample Point

Write the Sample Number and
the Density at that Point

②

$\left(\begin{array}{c}2\end{array}\right)$
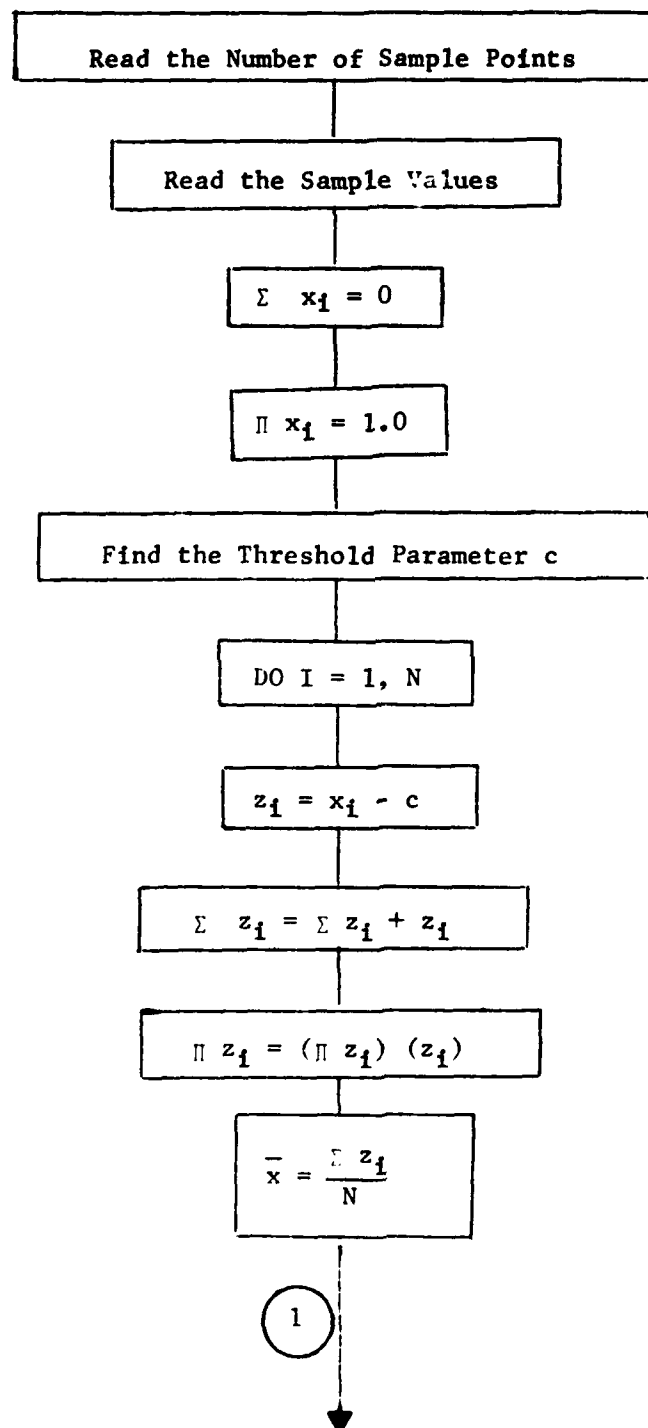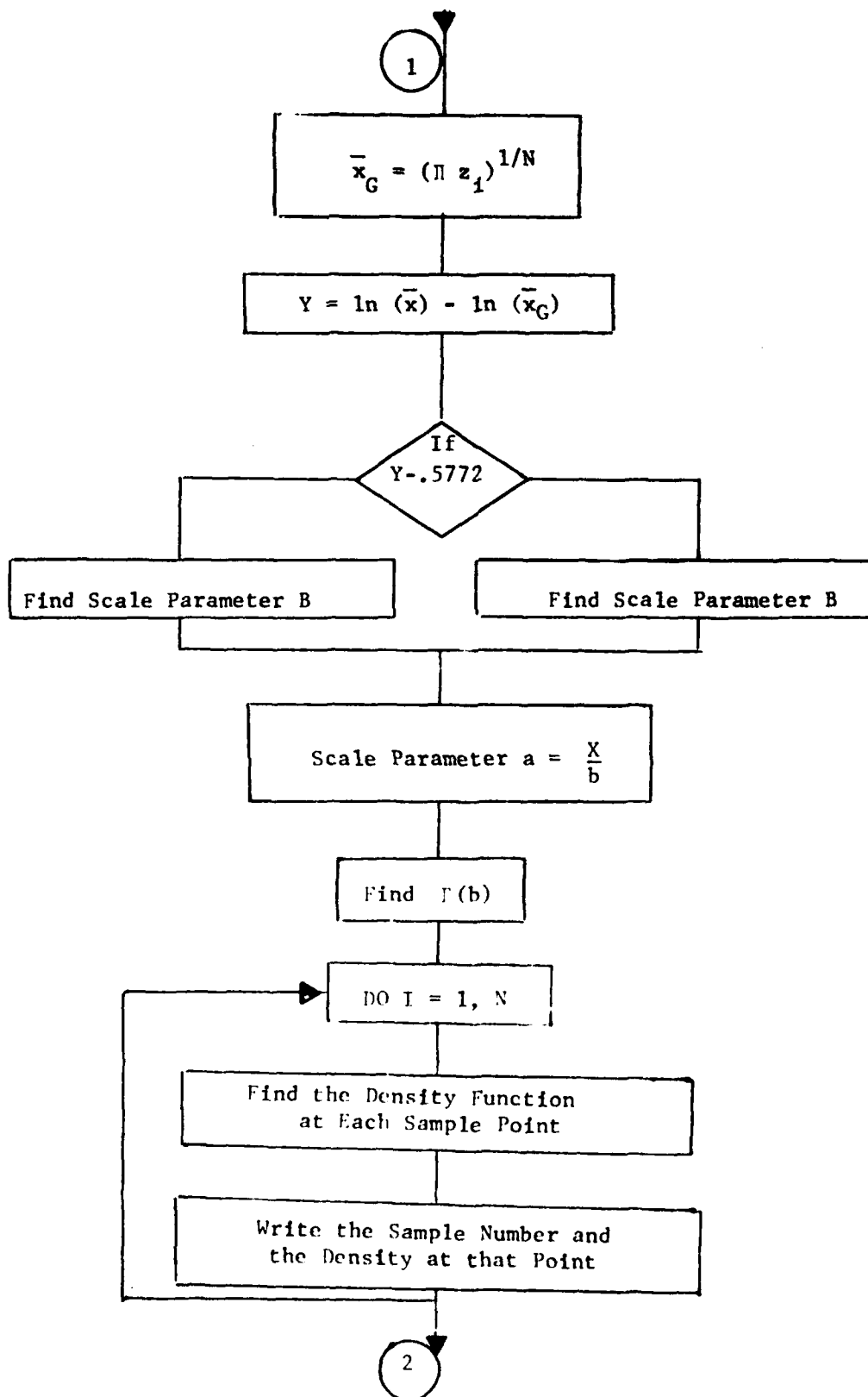
Write the Values of the
Threshold, Scale and Shape Parameters

```
      DIMENSION X(100),Z(100)
      READ(5,10) N
      READ(5,20) (X(M), M=1,N)
C     THERE ARE N SAMPLE POINTS X(1),X(2),...X(N)
      SXBAR = 0.0
      SXDOT = 1.0
      AN = N
      C=(X(2)+X(1))/2.0-(X(3)-X(1))*(X(3)-X(2))/(4.*X(2)-2.
     1 *X(1)-2.*X(3))
C     C IS THE THRESHOLD PARAMETER
      DO 100 L = 1, N
      Z(L) = X(L)-C
      SXBAR - SXBAR + Z(L)
      SXDOT = SXDOT * Z(L)**(1.0/AN)
  100 CONTINUE
      XBAR = SXBAR/AN
      Y = ALOG(XBAR/SXDOT)
C     XBAR IS THE ARITHMETIC MEAN, SXDOT THE GEOMETRIC MEAN
      IF(Y-.5772) 1,1,2
    1 B = (.5000876+.1648852*Y-.0544274*Y*Y)/Y
      GO TO 3
    2 B = (8.898919+9.05995*Y+.9775373*Y*Y)/(Y*(17.79728
     1 + 11.968477 * Y + Y*Y))
C     B IS THE SHAPE PARAMETER
    3 A = XBAR/B
C     A IS THE SCALE PARAMETER
C     VALUE OF THE DENSITY FUNCTION AT EACH SAMPLE POINT.
C     THE NEXT SIXTEEN CARDS ARE OPTIONAL AND WRITE THE
C     THE FIRST SEVEN CARDS FIND GAMMA(B)
      WRITE(6,30)
   30 FORMAT(1H1,14X,'SAMPLE',8X,'SAMPLE',13X,'DENSITY'/
     1 14X,'NUMBER',9X,'POINT'//)
      AMUL = 1.0
      BG=B
    7 IF(BG - 2.0) 4,4,5
    5 BG = BG - 1.0
      AMUL = AMUL * BG
      GO TO 7
    4 Q = BG - 1.0
      GAMB=AMUL*(1.-.57710166*Q+.98585399*Q**2.-.87642128*Q
     1 **3.+.8328212*Q**4.-.5684729*Q**5.+.25482049*Q**6.-
     2 .0514993*Q**7.)
      DO 200 K = 1,N
      FX = ((Z(K)/A)**(B-1.)*EXP((-1.)*Z(K)/A))/(A*GAMB)
      WRITE(6,40) K, X(K), FX
   40 FORMAT(15X,I3,10X,F6.2,5X,F16.7)
  200 CONTINUE
      WRITE(6,50) C, A, B
   10 FORMAT(I3)
   20 FORMAT(12F6.2)
```

```
50    FORMAT(1HO,'THE DATA IS GAMMA DISTRIBUTED WITH ',
     1'THRESHOLD PARAMETER = ',F7.4//40X,'SCALE PARAMETER = '
     2 ,F8.4//36X,'AND SHAPE PARAMETER = ',F8.4)
      STOP
      END
```

CHAPTER III

THE GAMMA DISTRIBUTION

In this chapter, the gamma distribution will be investigated.
The moments and the moment generating function along with the
convolution of the gamma distribution will be given and the method
of parameter estimation will be developed.

The gamma density function is given by

$$f(t) = \begin{cases} \dfrac{\left(\dfrac{t-c}{a}\right)^{b-1} e^{-\frac{t-c}{a}}}{a\ \Gamma(b)} & c < t < \infty \\ 0 & \text{otherwise} \end{cases}$$

where c is the threshold value,

  a is the scale parameter,

  b is the shape parameter, and

  $\Gamma(n)$ is the gamma function with argument n (Table 2).

The cummulative distribution is

$$F(t_0) = Pr\ (t < t_0) = I\ (u,p)$$

where I (u,p) is the Incomplete Gamma Function Ratio and in Harter's
notation (5)

$$p = b - 1$$

and

$$u = \frac{t_0 - c}{a\sqrt{b-1}}\ .$$

The moments and the variance are found to be

35

SAMPLE NUMBER	SAMPLE POINT	DENSITY
1	0.58	0.0659849
2	0.89	0.2518429
3	0.10	0.3828394
4	1.27	0.4624169
5	1.42	0.5057862
6	1.55	0.5227300
7	1.67	0.5229904
8	1.78	0.5122156
9	1.89	0.4929394
10	2.00	0.4671661
11	2.11	0.4367902
12	2.22	0.4035086
13	2.33	0.3687756
14	2.45	0.3306277
15	2.58	0.2903473
16	2.73	0.2466032
17	2.90	0.2018121
18	3.11	0.1544704
19	3.42	0.1006275

THE DATA IS GAMMA DISTRIBUTED WITH THRESHOLD PARAMETER = 0.1890

SCALE PARAMETER = 0.3886

AND SHAPE PARAMETER = 4.6609

INPUT FORMAT COLUMN IDENTIFICATION

NUMBER OF SAMPLE POINTS, N, FORMAT   (I3) <u>0</u> <u>1</u> <u>9</u>

NUMBER OF SAMPLE POINTS

INPUT FORMAT COLUMN IDENTIFICATION

SAMPLE VALUES, X (I)

FORMAT (12F6.2) _ _ 1 . 3 6

SAMPLE VALUES

APPENDIX 2

Computer Program for Convolving

Several Sets of Data

This program utilizes the previous program in a do loop to fit

gamma distributions to several sets of data. The average scale para-

meter is found and each set of data is forced to have this common

scale parameter. The distributions are convolved with the threshold

parameter being $\sum_{i=1}^{k} c_i$, scale parameter $= \dfrac{\sum_{i=1}^{k} a_i}{k}$ and shape parameter $=$

$\sum_{i=1}^{k} b_i$ where there are k sets of data.

FLOWCHART

```
  ┌─────────────────────────────────────────┐
  │   Read the Number of Subsystems, K       │
  └─────────────────────────────────────────┘
                      │
            ┌──────────────────┐
            │  $\Sigma\ c_i = 0$  │
            └──────────────────┘
                      │
            ┌──────────────────┐
            │  $\Sigma\ a_i = 0$  │
            └──────────────────┘
                      │
            ┌──────────────────┐
            │  $\Sigma\ b_i = 0$  │
            └──────────────────┘
                      │
            ┌──────────────────┐
    ┌──────▶│   DO I = 1, K     │
    │       └──────────────────┘
    │                 │
    │     ┌────────────────────────────┐
    │     │  Read the Number of Sample │
    │     │   Points in the Subsystem  │
    │     └────────────────────────────┘
    │                 │
    │     ┌────────────────────────────┐
    │     │   Read the Sample Values   │
    │     │     for the Subsystem      │
    │     └────────────────────────────┘
    │                 │
    │     ┌────────────────────────────┐
    │     │   Fit a Gamma Distribution │
    │     │      as in Program 1       │
    │     └────────────────────────────┘
    │                 │
    │     ┌────────────────────────────┐
    │     │  $a_i = \Sigma\ a_i + a_i$   │
    │     └────────────────────────────┘
    │                 │
    │     ┌────────────────────────────┐
    │     │  $c_i = \Sigma\ c_i + c_i$   │
    │     └────────────────────────────┘
    │                 │
   (1)              (2)
    │                 │
    ▼                 ▼
```
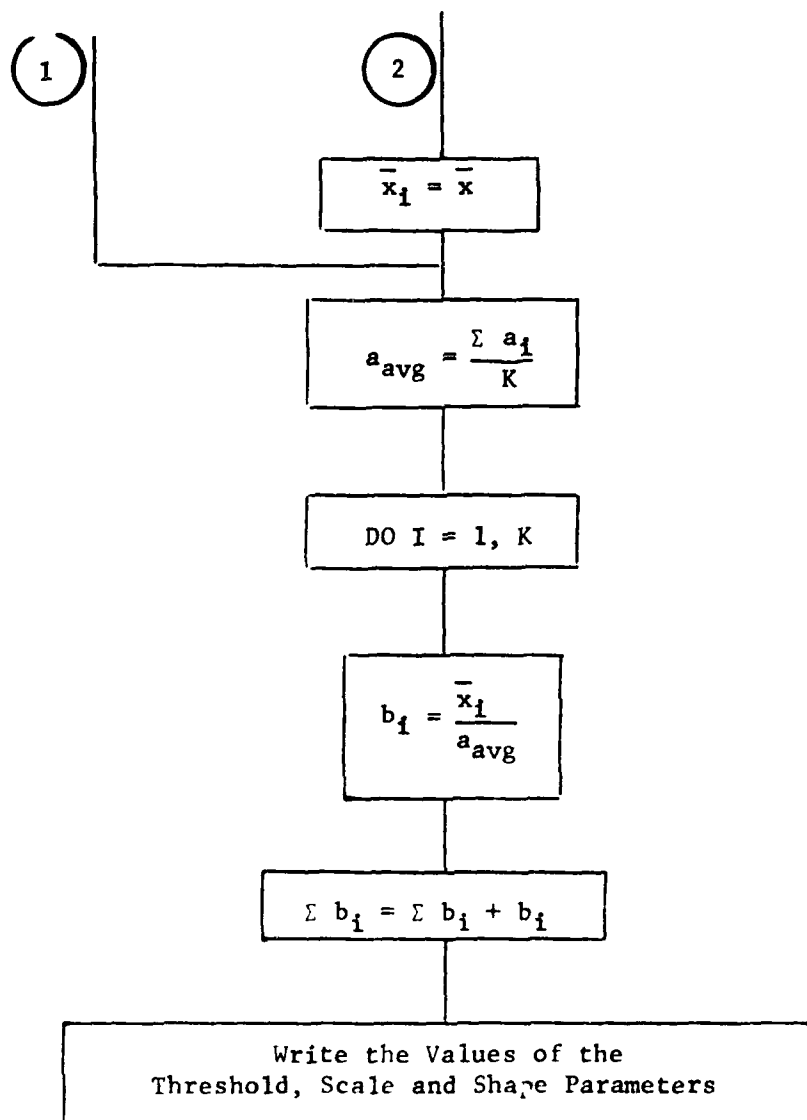
① ②

$$\overline{x}_i = \overline{x}$$

$$a_{avg} = \frac{\Sigma\ a_i}{K}$$

$$DO\ I = 1,\ K$$

$$b_i = \frac{\overline{x}_i}{a_{avg}}$$

$$\Sigma\ b_i = \Sigma\ b_i + b_i$$

Write the Values of the
Threshold, Scale and Shape Parameters
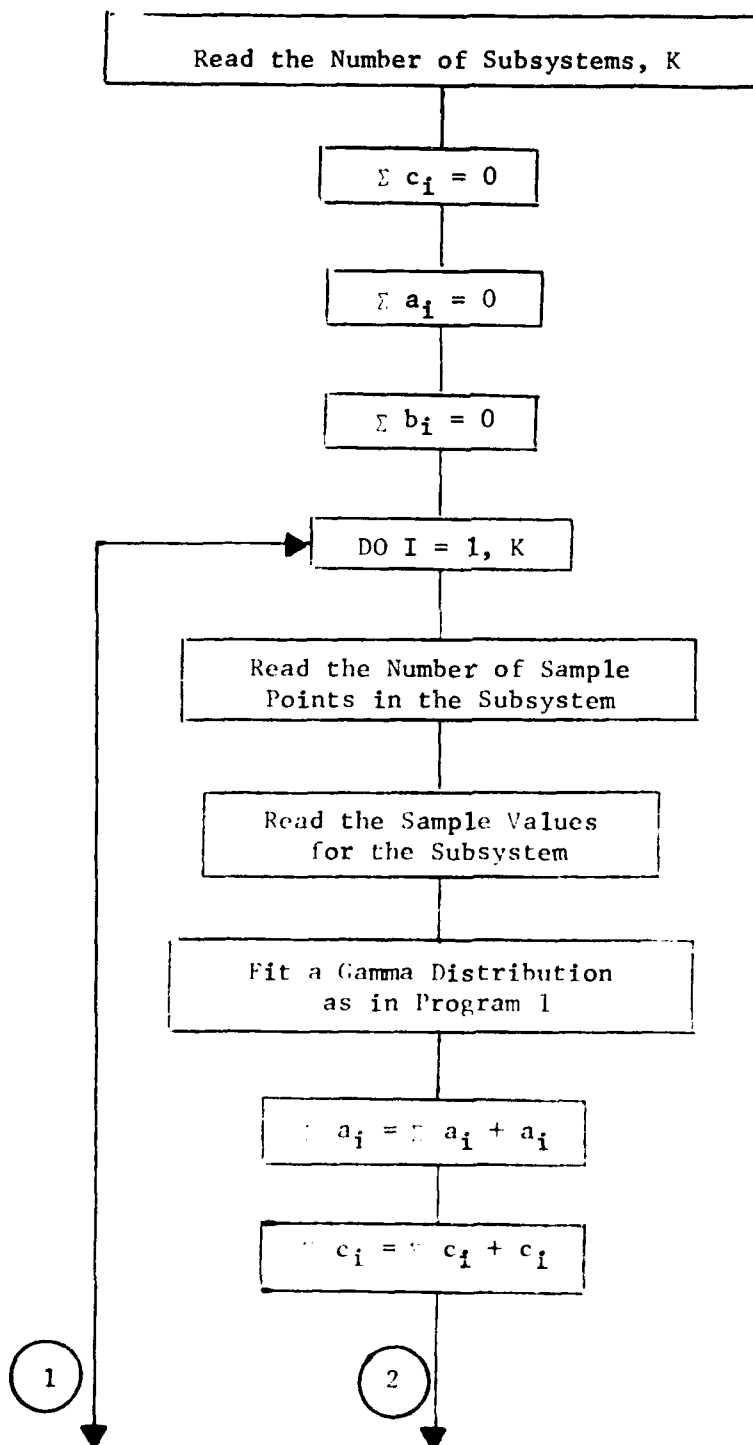
```
          DIMENSION X(100)
          READ(5,10) K
C         THERE ARE K SUB-SYSTEMS BEING ANALYZED
          SUMC = 0.0
          SUMA = 0.0
          SUMB = 0.0
C         THE DO 800 LOOP FITS A GAMMA DISTRIBUTION TO EACH SET
C         OF DATA.
          DO 800 I = 1,K
          READ(5,10) N
          READ(5,20) (X(M), M=1,N)
C         THERE ARE N SAMPLE POINTS X(1),X(2),...X(N)
          SXBAR = 0.0
          SXDOT = 1.0
          AN = N
          C=(X(2)+X(1))/2.0-(X(3)-X(1))*)X(3)-X(2))/(4.*X(2)
         1 - 2.*X(1) - 2.*X(3))
C         C IS THE THRESHOLD PARAMETER
          DO 100 L = 1, N
          Z(L) = X(L)-C
          SXBAR = SXBAR + Z(L)
          SXDOT = SXDOT * Z(L)**(1.0/AN)
  100     CONTINUE
          XBAR = SXBAR/AN
          XB(I) = XBAR
          Y = ALOG(XBAR/SXDOT)
C         XBAR IS THE ARITHMETIC MEAN, SXDOT THE GEOMETRIC MEAN
          IF(Y-.5772) 1,1,2
    1     B = (.5000876+.1648852*Y-.0544274*Y*Y)/Y
          GO TO 3
    2     B = (8.898919+9.05995*Y+.9775373*Y*Y)/(Y*(17.79728
         1 + 11.968477 * Y + Y*Y))
C         B IS THE SHAPE PARAMETER
    3     A = XBAR/B
C         A IS THE SCALE PARAMETER
          SUMA = SUMA + A
          SUMC = SUMC + C
C         THIS WRITE STATEMENT IS OPTIONAL AND PRINTS THE
C         PARAMETERS OF EACH SUB-SYSTEM.
          WRITE(6,30) I, C, A, B
   30     FORMAT(1H0,'SUB-SYSTEM',I3,'HAS BEEN FITTED WITH A',
         1' GAMMA DISTRIBUTION AND'//' C =',F8.4,' , A =',F8.4,
         2' AND B = ',F8.4)
  800     CONTINUE
          AK = K
          AAVG = SUMA/AK
          DO 900 L = 1,K
          SUMB = SUMB + XB(L)/AAVG
  900     CONTINUE
          WRITE(6,40) K, SUMC, AAVG, SUMB
   10     FORMAT(I3)
```

```
20    FORMAT(12F6.2)
40    FORMAT(1H0,I3,' SUBSYSTEMS HAVE BEEN FITTED WITH GAMMA'
     1,' DISTRIBUTIONS AND THEIR CONVOLUTION IS'//' A GAMMA'
     2,' DISTRIBUTION WITH C = ',F8.4,' A =',F8.4,' AND B ='
     3F8.4)
      STOP
```

SUB-SYSTEM 1 HAS BEEN FITTED WITH A GAMMA DISTRIBUTION AND

C = 0.9400 , A = 0.4600 AND B = 4.4778

SUB-SYSTEM 2 HAS BEEN FITTED WITH A GAMMA DISTRIBUTION AND

C = 2.6793 , A = 0.5364 AND B = 4.3263

SUB-SYSTEM 3 HAS BEEN FITTED WITH A GAMMA DISTRIBUTION AND

C = 2.0327 , A = 0.03400 AND B = 11.6674

SUB-SYSTEM 4 HAS BEEN FITTED WITH A GAMMA DISTRIBUTION AND

C = 0.1890 , A = 0.3886 AND B = 4.6609

4 SUBSYSTEMS HAVE BEEN FITTED WITH GAMMA DISTRIBUTIONS AND THEIR CONVOLUTION IS

A GAMMA DISTRIBUTION WITH C = 5.8410 A = 0.4313 AND B - 23.5563

44

INPUT FORMAT COLUMN IDENTIFICATION

NUMBER OF SUBSYSTEMS, K, FORMAT (I3) 0 0 4

NUMBER OF SUBSYSTEMS

INPUT FORMAT COLUMN IDENTIFICATION

NUMBER OF SAMPLE POINTS, N, FORMAT (I3) 0 1 9

NUMBER OF SAMPLE POINTS

INPUT FORMAT COLUMN IDENTIFICATION

SAMPLE VALUES, X (I)

FORMAT (12F6.2) _ _ 1 . 3 6

SAMPLE VALUES

APPENDIX 3

Derivation of the Equation

for the Threshold Parameter

As indicated in Chapter III, the maximum likelihood equation for the threshold parameter requires an iteration process. As this parameter is rarely known in maintainability analysis, some estimate of this parameter must be made from the data available. Since the area between adjacent ordered sample points is known from order statistics to be $\frac{1}{n+1}$ (10), a straight line can be fitted to the unknown curve and the threshold parameter estimated.
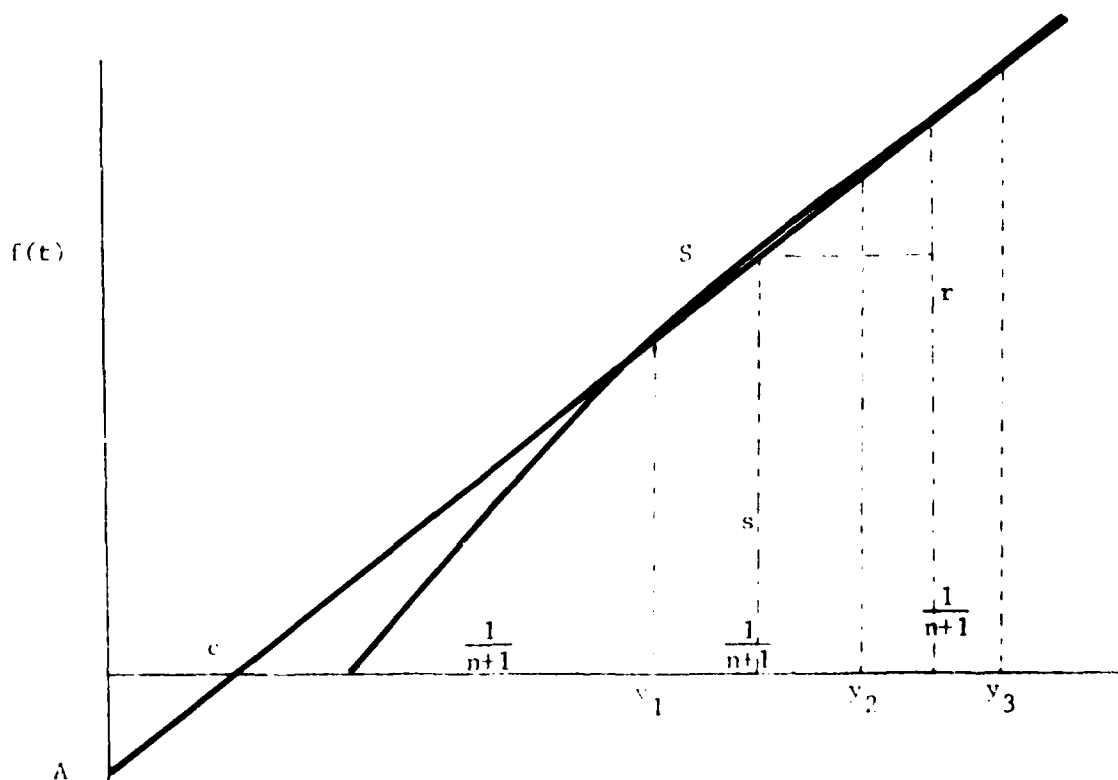
The area between the first and second sample points is

$$\frac{1}{n+1} = s(y_2 - y_1)$$

and the area between the second and third points is

$$\frac{1}{n+1} = r(y_3 - y_2).$$

The slope of the line is then

$$\text{slope} = \frac{2(r - s)}{y_3 - y_1}$$

and at the point S

$$s = A + \text{slope} \; \frac{y_2 + y_1}{2}$$

$$A = s - \text{slope} \; \frac{y_2 + y_1}{2} . \qquad\qquad \text{Equation 4}$$

At the point c

$$0 = A + \text{slope} \; (c)$$

$$A = - \text{slope} \; (c). \qquad\qquad \text{Equation 5}$$

Equating equation 4 to equation 5

$$s - \text{slope} \; \frac{y_2 + y_1}{2} = - \text{slope} \; (c)$$

and

$$c = \frac{y_2 + y_1}{2} = - \frac{s}{\text{slope}}$$

$$= \frac{y_2 + y_1}{2} - \frac{(y_3 - y_2)(y_2 - y_1)}{2(2y_2 - y_1 - y_3)}$$

LIST OF REFERENCES

1. Brockmeyer, E., Halstrom, H. L. and Jensen, A., The Life and Works of A. K. Erlang, Trans. Danish Acad. Sci., 1948.

2. Chapman, D. G., "Estimating the Parameters of a Truncated Gamma Distribution", Ann. Math. Stat., Vol. 27, No. 2, June 1956.

3. Collins, C. A., thesis for a Master's of Science majoring in Statistics, Texas A&M University, 1966 (unpublished).

4. Davis, H. T., Tables of Higher Mathematical Functions, Vol. 1, Principia Press, 1933.

5. Greenwood, J. A. and Durand, D., "Aids for Fitting the Gamma Distribution by Maximum Likelihood", Technometrics, Vol. 2, No. 1, February 1960.

6. Harter, H. L. and Moore, A. H., "Maximum Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and from Censored Samples", Technometrics, Vol. 7, No. 4, November 1965.

7. Harter, H. L., "Maximum-Likelihood Estimation of the Parameters of a Four Parameter Generalized Gamma Distribution from Complete and Censored Samples", Technometrics, Vol. 9, No. 1, February 1967.

8. Harter, H. L., Tables of the Incomplete Gamma Function Ratio and of Percentage Points of the Chi-Square and Beta Distributions, U. S. Government Printing Office, Washington, D. C., 1965, Price-$2.50.

9. Hogg, R. V. and Craig, A. T., Introduction to Matehmatical Statistics, Macmillan Company, New York, New York, 1959.

10. Johnson, N. L. and Leone, F. C., Statistics and Experimental Design in Engineering and the Physical Sciences, Vol. 1, John Wiley and Sons, Inc., New York, New York, 1964.

11. Ostle, B., Statistics in Research, Iowa State University Press, Ames, Iowa, 1963.

12. Farr, V. R., and Webster, J. T., "A Method for Discriminating Between Failure Density Functions Used in Reliability Predictions", Technometrics, Vol. 7, No. 1, February 1965.

13. Pearson, K., Tables of the Incomplete Gamma Function, Cambridge University Press, 1934.

14. Petterson, E. L., "Maintainability Risk Analysis Using the Analytical Maintenance Model", 1967.

15. Pieruschka, E., _Principals of Reliability_, Prentice-Hall, Englewood Cliffs, New Jersey, 1963.

16. Stacy, E. W., "A Generalization of the Gamma Distribution", _Ann. Math. Stat._, Vol. 33, No. 3, September 1962.

17. Stacy, E. W. and Mihram, G. A., "Parameter Estimation for a Generalized Gamma Distribution", _Technometrics_, Vol. 7, No. 3, August 1965.

CHAPTER III

THE GAMMA DISTRIBUTION

In this chapter, the gamma distribution will be investigated. The moments and the moment generating function along with the convolution of the gamma distribution will be given and the method of parameter estimation will be developed.

The gamma density function is given by

$$f(t) = \begin{cases} \dfrac{\left(\dfrac{t-c}{a}\right)^{b-1} e^{-\frac{t-c}{a}}}{a\ \Gamma(b)} & c < t < \infty \\ \\ 0 & \text{otherwise} \end{cases}$$

where c is the threshold value,

    a is the scale parameter,

    b is the shape parameter, and

    $\Gamma(n)$ is the gamma function with argument n (Table 2).

The cummulative distribution is

$$F(t_0) = Pr\ (t < t_0) = I\ (u,p)$$

where I (u,p) is the Incomplete Gamma Function Ratio and in Harter's notation (8)

$$p = b - 1$$

and

$$u = \frac{t_0 - c}{a\sqrt{b-1}}.$$

The means and variances are then found to be